MULTIVARIATE STATISTICAL ANALYSIS OF PHYTOPLANKTON POPULATIONS IN THE LOWER HUDSON RIVER

Kim, E-W[1]; **Park**, M. [2]; **Levandowsky**, M. [3]; **Wang**, Zhaoyan[4]; **Vaccari**, David A[4]

An 8-year series of weekly phytoplankton samples from surface water at two lower Manhattan sites in the Hudson River was analyzed using Multivariate Polynomial Regression (MPR) and Canonical Correspondence Analysis (CCA). The 2 sites were Pier 26 on the Hudson River and Piers 15-16 on the East River. Presence-absence data were collected on 29 taxa, readily identifiable with the light microscope from living material, along with measurements of temperature, salinity, pH, Secchi depth and dissolved oxygen.

The CCA analysis used presence-absence data of all 29 taxa at both sites. The Hudson sample points formed distinct clusters with respect to salinity and temperature, whereas the East River samples, involving the same taxa, did not. The data suggest significant differences in organization of phytoplankton at these two sites: the lower Hudson River phytoplankton community appears to be much more structured than that of the East River. This may be due to hydrodynamic differences: the Hudson is typically highly stratified, whereas the East River is dominated by turbulent mixing. The contrasts in the data appear to reflect the high degree of hydrologic stratification of lower Hudson, and the extreme turbulence of the East River. Distributions of individual taxa in the ordination plane were also studied, allowing a characterization of them with regard to hydrographic variables.

MPR models were developed to predict the probability of presence or absence of 3 of the phytoplankton taxa at the Hudson River site: *Actinoptychus undulata, Rhizosolenia setigera*, and *Scenedesmus quadricauda* from the physicochemical measurements. The MPR model can describe interactions and nonlinear behavior in complex systems with an explicit equation. The MPR model significantly improved the accuracy of prediction over multivariate linear models, and could describe complex behavior such as nonlinearities and interactions. A probability response surface can be described using the model. This was used to identify the physical parameter space corresponding to various probabilities of occurrence for the taxa that were modeled.

These analyses demonstrate that meaningful statistical information can be obtained from such categorical (presence/absence) data. This information includes identifying correlations between sets of population and water quality variables, and being able to predict the probability of finding a particular species given independent physical and chemical water parameters.

[1] Haskins Labs, Pace University, N.Y., NY
[2] The River Project, N.Y., NY
[3] Haskins Labs, Pace University, N.Y., NY, and The River Project, N.Y., NY
[4] Dept. of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ